WHAT IS CLAIMED IS:

1. A method, comprising:

receiving a request to data at a memory address in a first memory device, wherein data in the first memory device is cached in a second memory device;

determining whether to fetch the requested data from the first memory device to cache in the second memory device in response to determining that the requested data is not in the second memory device; and

accessing the requested data in the first memory device and bypassing the second memory device to execute the request in response to determining not to fetch the requested data from the first memory device to cache in the second memory device.

2. The method of claim 1, wherein the memory address to which the request is directed comprises a first memory address, wherein the determination to fetch the requested data from the first memory device is made in response to additionally determining that a location in the second memory device into which the fetched requested data from the first memory device is cached includes updated data for the first memory device.

3. The method of claim 1, wherein the request may comprise a request to read or write data at the memory address in the first memory device.

4. The method of claim 1, further comprising:

fetching the requested data from the memory address in the first memory device to cache in the second memory device in response to determining to fetch the requested data from the first memory device to the second memory device as part of executing the request.

5. The method of claim 1, wherein determining whether to fetch the requested data from the first memory device to the second memory device in response to determining that the requested data is not in the cache comprises:

determining whether an amount of valid data from the first memory device cached in the second memory device exceeds a threshold.

6. The method of claim 5, wherein the requested data is fetched from the first memory device to the second memory device in response to determining that the amount of valid data cached in the second memory device is below the threshold and wherein the requested data is accessed in the first memory device and the second memory device bypassed when executing the request in response to determining that the amount of valid data cached in the second memory device exceeds the threshold.

7. The method of claim 5, further comprising:

incrementing a counter whenever data is fetched from the first memory device and cached in the second memory device; and

decrementing the counter whenever data is invalidated in the second memory device, wherein determining whether the amount of valid data from the first memory device being cached in the second memory device exceeds the threshold comprises determining whether the counter exceeds the threshold and determining whether the amount of valid data in the second memory device is below the threshold comprises determining whether the counter is below the threshold.

8. The method of claim 7, wherein the second memory device includes cache lines for storing data from memory addresses in the first memory device, wherein the counter is incremented when data from the first memory device is fetched and written to one cache line in the second memory device and wherein the counter is decremented when data is removed from one cache line in the second memory device.

9. The method of claim 8, wherein there is one valid flag for each cache line indicating whether the cache line includes valid data from the first memory device, wherein the counter is incremented when the valid flag is changed from indicating invalid to indicating valid data and decremented when the valid flag is changed from indicating valid to indicating invalid data.

10. The method of claim 1, wherein the first memory device is a slower access memory device than the second memory device.

11. The method of claim 10, wherein the second memory device comprises a cache embedded in a processing unit and wherein the first memory device is external to the processing unit.

12. A system, comprising:

a first memory device;

a second memory device;

circuitry, wherein the circuitry is operable to:

(i) receive a request to data at a memory address in a first memory device, wherein data in the first memory device is cached in a second memory device;

(ii) determine whether to fetch the requested data from the first memory device to cache in the second memory device in response to determining that the requested data is not in the second memory device; and

(iii) access the requested data in the first memory device and bypassing the second memory device to execute the request in response to determining not to fetch the requested data from the first memory device to cache in the second memory device.

13. The system of claim 12, wherein the memory address to which the request is directed comprises a first memory address, wherein the determination to fetch the requested data from the first memory device is made in response to additionally determining that a location in the second memory device into which the fetched requested data from the first memory device is cached includes updated data for the first memory device.

14. The system of claim 12, wherein the request may comprise a request to read or write data at the memory address in the first memory device.

15. The system of claim 12, wherein the circuitry is further operable to:

fetch the requested data from the memory address in the first memory device to cache in the second memory device in response to determining to fetch the requested data from the first memory device to the second memory device as part of executing the request.

16. The system of claim 12, wherein determining whether to fetch the requested data from the first memory device to the second memory device in response to determining that the requested data is not in the cache comprises:

determine whether an amount of valid data from the first memory device cached in the second memory device exceeds a threshold.

17. The system of claim 16, wherein the requested data is fetched from the first memory device to the second memory device in response to determining that the amount of valid data cached in the second memory device is below the threshold and wherein the requested data is accessed in the first memory device and the second memory device bypassed when executing the request in response to determining that the amount of valid data cached in the second memory device exceeds the threshold.

18. The system of claim 16, further comprising:

a counter, wherein the circuitry is further operable to:

(i) increment a counter whenever data is fetched from the first memory device and cached in the second memory device; and

(ii) decrement the counter whenever data is invalidated in the second memory device, wherein determining whether the amount of valid data from the first memory device being cached in the second memory device exceeds the threshold comprises determining whether the counter exceeds the threshold and determining whether the amount of valid data in the second memory device is below the threshold comprises determining whether the counter is below the threshold.

19. The system of claim 18, wherein the second memory device includes cache lines for storing data from memory addresses in the first memory device, wherein the counter is incremented when data from the first memory device is fetched and written to one cache line in the second memory device and wherein the counter is decremented when data is removed from one cache line in the second memory device.

20. The system of claim 19, further comprising:

one valid flag for each cache line indicating whether the cache line includes valid data from the first memory device, wherein the counter is incremented when the valid flag is changed from indicating invalid to indicating valid data and decremented when the valid flag is changed from indicating valid to indicating invalid data.

21. The system of claim 12, wherein the first memory device is a slower access memory device than the second memory device.

22. The system of claim 21, wherein the second memory device comprises a cache embedded in a processing unit and wherein the first memory device is external to the processing unit.

23. A system, comprising:

processor implemented on a first integrated circuit device;

a first memory device implemented on a second integrated circuit device;

a second memory device implemented on the first integrated circuit device;

circuitry, wherein the circuitry is operable to:

(i) receive a request to data at a memory address in a first memory device, wherein data in the first memory device is cached in a second memory device;

(ii) determine whether to fetch the requested data from the first memory device to cache in the second memory device in response to determining that the requested data is not in the second memory device; and

(iii) access the requested data in the first memory device and bypassing the second memory device to execute the request in response to determining not to

fetch the requested data from the first memory device to cache in the second memory device.

24. The system of claim 23, wherein the memory address to which the request is directed comprises a first memory address, wherein the determination to fetch the requested data from the first memory device is made in response to additionally determining that a location in the second memory device into which the fetched requested data from the first memory device is cached includes updated data for the first memory device.

25. An article of manufacture for managing data in a first memory device and a second memory device, wherein the article of manufacture is operable to:

receive a request to data at a memory address in a first memory device, wherein data in the first memory device is cached in a second memory device;

determine whether to fetch the requested data from the first memory device to cache in the second memory device in response to determining that the requested data is not in the second memory device; and

access the requested data in the first memory device and bypassing the second memory device to execute the request in response to determining not to fetch the requested data from the first memory device to cache in the second memory device.

26. The article of manufacture of claim 25, wherein the memory address to which the request is directed comprises a first memory address, wherein the determination to fetch the requested data from the first memory device is made in response to additionally determining that a location in the second memory device into which the fetched requested data from the first memory device is cached includes updated data for the first memory device.

27. The article of manufacture of claim 25, wherein the request may comprise a request to read or write data at the memory address in the first memory device.

28. The article of manufacture of claim 25, wherein the article of manufacture is further operable to:

fetch the requested data from the memory address in the first memory device to cache in the second memory device in response to determining to fetch the requested data from the first memory device to the second memory device as part of executing the request.

29. The article of manufacture of claim 25, wherein determining whether to fetch the requested data from the first memory device to the second memory device in response to determining that the requested data is not in the cache comprises:

determining whether an amount of valid data from the first memory device cached in the second memory device exceeds a threshold.

30. The article of manufacture of claim 29, wherein the requested data is fetched from the first memory device to the second memory device in response to determining that the amount of valid data cached in the second memory device is below the threshold and wherein the requested data is accessed in the first memory device and the second memory device bypassed when executing the request in response to determining that the amount of valid data cached in the second memory device exceeds the threshold.

31. The article of manufacture of claim 29, wherein the article of manufacture is further operable to:

increment a counter whenever data is fetched from the first memory device and cached in the second memory device; and

decrement the counter whenever data is invalidated in the second memory device, wherein determining whether the amount of valid data from the first memory device being cached in the second memory device exceeds the threshold comprises determining whether the counter exceeds the threshold and determining whether the amount of valid data in the second memory device is below the threshold comprises determining whether the counter is below the threshold.

32. The article of manufacture of claim 31, wherein the second memory device includes cache lines for storing data from memory addresses in the first memory device, wherein the counter is incremented when data from the first memory device is fetched and written to one cache line in the second memory device and wherein the counter is decremented when data is removed from one cache line in the second memory device.

33. The article of manufacture of claim 32, wherein there is one valid flag for each cache line indicating whether the cache line includes valid data from the first memory device, wherein the counter is incremented when the valid flag is changed from indicating invalid to indicating valid data and decremented when the valid flag is changed from indicating valid to indicating invalid data.

34. The article of manufacture of claim 25, wherein the first memory device is a slower access memory device than the second memory device.

35. The article of manufacture of claim 34, wherein the second memory device comprises a cache embedded in a processing unit and wherein the first memory device is external to the processing unit.